

## 日 本 国 特 許 庁

PATENT OFFICE  
JAPANESE GOVERNMENT

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出 願 年 月 日

Date of Application:

2000年 1月31日

出 願 番 号

Application Number:

特願2000-027201

出 願 人

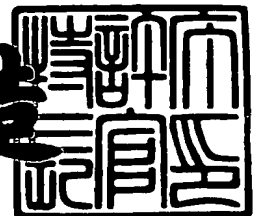
Applicant (s):

株式会社日立製作所

2000年12月 8日

特許庁長官  
Commissioner,  
Patent Office

及 川 耕 造



出証番号 出証特2000-3102000

【書類名】 特許願

【整理番号】 K00000171

【提出日】 平成12年 1月31日

【あて先】 特許庁長官殿

【国際特許分類】 G06F 11/20

【請求項の数】 4

【発明者】

    【住所又は居所】 神奈川県小田原市国府津 2 8 8 0 番地 株式会社日立製作所 ストレージシステム事業部内

    【氏名】 田中 幸一

【発明者】

    【住所又は居所】 神奈川県小田原市国府津 2 8 8 0 番地 株式会社日立製作所 ストレージシステム事業部内

    【氏名】 ▲高▼本 賢一

【特許出願人】

    【識別番号】 000005108

    【氏名又は名称】 株式会社日立製作所

【代理人】

    【識別番号】 100075096

    【弁理士】

    【氏名又は名称】 作田 康夫

【手数料の表示】

    【予納台帳番号】 013088

    【納付金額】 21,000円

【提出物件の目録】

    【物件名】 明細書 1

    【物件名】 図面 1

    【物件名】 要約書 1

【プルーフの要否】 要

【書類名】 明細書

【発明の名称】 記憶制御装置

【特許請求の範囲】

【請求項 1】

複数の接続されたホストコンピュータの指示に基づいて記憶装置の制御を行う制御装置であって、

複数のポートを有し前記ホストコンピュータとのプロトコルを制御する独立したコントローラを前記ホストコンピュータ毎に備え、このコントローラの障害時には障害を起こしたコントローラと接続された前記ホストコンピュータの接続を他のコントローラに切り替える制御装置。

【請求項 2】

前記切り替え時には、前記他のコントローラは前記障害を起こしたコントローラの情報を引き継ぐ請求項 1 に記載の制御装置。

【請求項 3】

前記障害を起こしたコントローラの障害回復時には、前記他のコントローラから制御を戻す請求項 1 に記載の制御装置。

【請求項 4】

前記制御を戻す契機をユーザーが選択できる請求項 3 に記載の制御装置。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は、上位装置からの情報の入出力要求を制御する入出力制御装置に関し、特に上位装置とファイバチャネルにて接続された入出力制御装置に関する。

【0002】

【従来の技術】

コントローラおよびディスクなどの記憶装置に冗長性を持たせたものとしては、従来、2重の系で構成され、一方の系が現用系として稼動し、他方の系が待機系として稼動し、現用系障害時には、待機機系に処理を引き継ぐことによりデータ保全性の向上を図っていた。この待機系への引き継ぎは、障害発生側のポート

情報（WWN、ポートIDなど）も同時に引き継ぐことによりホスト側に意識させることなく切り替えを実現させている。

【 0 0 0 3 】

【発明が解決しようとする課題】

コンピュータシステムの大規模化、データ処理の高速化、データの大容量化に伴い、記憶制御装置に対する高性能、高信頼性、可用性の向上が強く望まれている。

【 0 0 0 4 】

従来技術では、信頼性向上のためホスト側にパス切り替えソフトウェアを有して障害時の系の切り替えを実現していた。また、記憶制御装置側の機能により2重化したコントローラを現用系／待機系に分け制御していた。

【 0 0 0 5 】

本発明は、ホスト側にパス切り替えソフトウェアを使用することなく、2重化している全ての構成要素を有効に使用し、現用系／待機系等区別することなく制御可能とし、一方の系にて障害が発生しても他方の系で継続制御することを目的とする。

【 0 0 0 6 】

【課題を解決するための手段】

上記目的を達成する為、ホストコンピュータと制御装置間のバスプロトコルを制御する独立した複数のポートを有する複数のコントローラ、各ポートの情報をコントローラ間で共用、自コントローラ及び自ポートにて障害を検出した場合自ポートをループから切り離す機能、他コントローラ及び他ポートで発生した障害を検出する機能、他ポートをループから切り離す機能、他ポートをループから切り離した後、自ポートをループに参加させる機能、障害部位の回復を検知する機能、障害部位の回復を検知後、自ポートをループから切り離す機能、障害部位の回復時ループに参加する機能を設ける。

【 0 0 0 7 】

【発明の実施の形態】

以下、本発明の1実施例を図面を用いて説明する。

## 【 0 0 0 8 】

本実施例に示した構成は、あくまで1例であり、別の構成による実現も可能である。

## 【 0 0 0 9 】

図1は、本発明をディスクアレイ装置に適用した場合の1構成図である。

## 【 0 0 1 0 】

図1において、1001,1002はデータ処理を行う中央処理装置としてのホストコンピュータ、2000はディスクアレイ装置であり、各々のホストよりFibreチャネルバスで接続されている。

## 【 0 0 1 1 】

ディスクアレイ装置2000は、デュアル構成をとるコントローラ3000,4000、ホストコンピュータからのデータを格納するディスク装置群5000とから構成される。ディスク装置群5000は、アレイ構成をとることが多い。

## 【 0 0 1 2 】

コントローラ3000,4000は、ホストコンピュータ1001,1002とのプロトコル制御を行うホストI/F制御部3010,3020,4010,4020、データの転送を実行するデータ転送制御部3300、4300、ホストコンピュータ1001,1002とディスク装置群5000のデータ転送時に用いられる不揮発キャッシュ(以下、キャッシュと称す)3400,4400、各ディスク装置群5000とのプロトコル制御を行うDRV I/F制御部3500,4500より構成する。

## 【 0 0 1 3 】

又、各コントローラから、相手コントローラのホストI/F制御部には、専用線2100,2200,2300,2400が張られており、相手コントローラのホストI/F制御部をFibreループから切り離すことを可能としている。

## 【 0 0 1 4 】

ホストI/F制御部3010,4010は交替関係にある。本実施例では、3010をホストコンピュータ1001現用系として4010を待機系として設定している場合について説明する。同様にホストI/F制御部3020,4020は交替関係にあり、4020がホストコンピュータ1002の現用系に、3020が待機系に設定されている。本設定は一例であり、

他の設定でも、もちろんかまわない。

【 0 0 1 5 】

ホストI/F制御部3010、4010はFibreチャネルループ6000を介して、ホストコンピュータ1001に接続されている。正常時には、ホストI/F制御部3010がループ6000に参加しており、ホストI/F制御部4010は、待機状態にあり、ループ6000には参加していない。ここでは、プライベートループの例について示しているが、本発明は、Fibreチャネルの接続形態を限定するものではない。たとえば、接続形態がPoint-to-Point接続でも構わないし、Fabric装置を介しても構わない。

【 0 0 1 6 】

同様に、ホストI/F制御部3020、4020はFibreチャネルループ7000を介して、ホストコンピュータ1002に接続されている。正常時には、ホストI/F制御部4020がループ7000に参加しており、ホストI/F制御部3020は、待機状態にあり、ループ7000には参加していない。

【 0 0 1 7 】

尚、上記のように本願発明では複数のコントローラ3000,4000からディスク装置群にアクセスが発生するので、各コントローラ間にて排他制御を行ない、同時にディスク装置群の同じ場所にアクセスが生じないようにしている。

【 0 0 1 8 】

次にキャッシュ3400,4400の内容について、図2を用いて説明する。

【 0 0 1 9 】

キャッシュ3400,4400は、データを格納するデータ部3410,4410とコントローラ間の通信に用いるコントローラ間通信エリア3420,4420とホストI/F制御情報(WWN、ポートIDなど)が格納されたI/F情報エリア3430,4430とからなる。

【 0 0 2 0 】

キャッシュ3400,4400はデータ転送制御3300,4300を介して、自コントローラ3000,4000上のキャッシュ3400,4400への読み／書きはもちろんのこと、相手コントローラ4000,3000上のキャッシュ4400,3400への読み／書きも行うことが可能である。

## 【 0 0 2 1 】

相手コントローラ4000,3000への書き込みを実施した場合には、通信線2500,2600を用いて書き込んだことを相手コントローラ4000,3000に通知する。本通知手段としては、通信線を使用しなくても相手コントローラ4000,3000で定期的にポーリングすることで代替可能である。

## 【 0 0 2 2 】

また、I/F情報エリア3430,4430には各ホストI/F制御部3010,3020,4010,4020毎にエリアがあり、各コントローラ3000,4000が常にI/Fの最新情報を両キャッシュ3400,4400に2重書きで格納している。

## 【 0 0 2 3 】

次に、本実施例における、ディスクサブシステムでの、ホストコンピュータ1001,1002からのI/O処理について図1を用いて説明する。まずホストコンピュータ1001からのI/O処理について説明する。

## 【 0 0 2 4 】

ホストコンピュータ1001からのI/O要求時、ループ6000に参加してしているホストI/F制御部3010が、この要求を受け取りコントローラ3000がI/O処理を実施する。ホストコンピュータ1001からの要求が読み込み要求であった場合には、要求データをディスク装置群5000より、DRVI/F制御部3500、データ転送制御部3300を介してキャッシュ3400に読み込み後、キャッシュ3400からデータ転送制御部3300、ホストI/F制御部3010を介してホストコンピュータ1001に転送される。

## 【 0 0 2 5 】

ホストコンピュータ1001からの要求が書き込み要求であった場合には、書き込みデータをホストI/F制御部3010、データ転送制御部3300を介してキャッシュ3400に書き込む。この際、データ転送制御部3300は、キャッシュ4400にも2重書きを行う。その後、データ転送制御部3300、DRVI/F制御部3500を介してディスク装置群5000にデータを書きこむ。

## 【 0 0 2 6 】

次に、ホストコンピュータ1002からのI/O処理について説明する。

【 0 0 2 7 】

ホストコンピュータ1002からのI/O要求時、ループ7000に参加してしているホストI/F制御部4020が、この要求を受け取りコントローラ4000がI/O処理を実施する。

【 0 0 2 8 】

読み込み処理、書きこみ処理は、前述のコントローラ3000の処理と同様である。

【 0 0 2 9 】

このように現用系のホストI/F制御部を両方のコントローラ3000,4000に割り振って設定すれば、正常時には、両方のコントローラを稼働させることが可能であり、一方のコントローラを現用系、もう一方を待機系として用いる従来技術に比べると、2コントローラ分の性能を引き出すことが可能であり、正常時、従来技術のほぼ2倍の性能向上を図ることができる。

【 0 0 3 0 】

次に、コントローラ3000またはホストI/F制御部3010に障害が発生した場合の処理方式について図3、図4を用いて説明する。

【 0 0 3 1 】

まず、上記障害をコントローラ3000が検知した場合について図3を用いて説明する。

【 0 0 3 2 】

コントローラ3000がコントローラ3000またはホストI/F制御部3010の障害を検知するとコントローラ3000は、まずホストI/F制御部3010をループから切り離し、次にコントローラ間通信エリア4420に障害情報をデータ転送制御部3300を介して書きこみ、コントローラ4000にホストI/F制御部3010の障害を通知する。この通知を受け取ったコントローラ4000は、I/F情報エリア4430からホストI/F制御部3010のI/F情報を取得し、この情報を用いて、ホストI/F制御部4010を待機状態から稼働状態に変更し、ループ6000に参加させる。

【 0 0 3 3 】

次に、コントローラ4000がコントローラ3000またはホストI/F制御部3010の障



害を検知した場合について図4を用いて説明する。

【0034】

コントローラ4000がコントローラ3000またはホストI/F制御部3010の障害を検知するとコントローラ4000は専用線2100を用いて、ホストI/F制御部3010をループ6000から切り離す。次にI/F情報エリア4430からホストI/F制御部3010のI/F情報を取得し、この情報を用いて、ホストI/F制御部4010を待機状態から稼動状態に変更し、ループ6000に参加させる。

【0035】

このようにどちらが障害を検知しても待機系のホストI/F制御部4010への切り替えが可能である。又、ホストI/F制御部3010の情報(WWN,ポートIDなど)を用いる為、ホストコンピュータ1001は、ホストI/F制御部が4010に変更となったことを全く意識することなくI/O処理の継続が可能である。このことにより、ホストコントローラ側にパス切り替え用のミドルウェアを組み込むことなしにパス切り替えが可能である。このパス切り替え用ミドルは一般に高価であり、本発明は、システムとしての低コスト化にもつながる。

【0036】

次に、コントローラ3000またはホストI/F制御部3010が障害から復旧した場合の処理方式について図5を用いて説明する。

【0037】

コントローラ3000は復旧が完了すると、コントローラ間通信エリア4420に復旧情報をデータ転送制御部3300を介して書きこみ、コントローラ4000にホストI/F制御部3010の復旧を通知する。コントローラ4000は、復旧を検知すると、ホストI/F制御部4010をループ6000から切り離し、切り離しが終了したことをコントローラ間通信エリア3420を通じてコントローラ3000に通知する。本通知を受けたコントローラ3000は、I/F情報エリア3430からホストI/F制御部4010のI/F情報を取得し、この情報を用いて、ホストI/F制御部3010を待機状態から稼動状態に変更し、ループ6000に参加させ、復旧を完了する。

【0038】

このように復旧時にも、ホストI/F制御部4010のI/F情報を使用する為、ホスト

コンピュータは意識することなく、復旧を完了させ、2コントローラを稼動状態に復旧させることができる。

【0039】

本実施例では、ホストI/F制御部3010のループへの復旧を障害復旧時に直に行う方式について示したが、復旧の契機は、ユーザが指定した契機でも構わないし、次回装置立ち上げ時でも構わない。又、これらの契機をユーザが選択できるようにしても構わない。

【0040】

【発明の効果】

本発明によれば、ホストコンピュータと制御装置間のバスプロトコルを制御する独立した複数のポートを有するコントローラを複数備え、正常時には、両方のコントローラを稼動させることが可能であり、一方のコントローラを現用系、もう一方を待機系として用いる従来技術に比べると、2コントローラ分の性能を引き出すことが可能であり、正常時、従来技術のほぼ2倍の性能向上を図ることができる。

【0041】

また、コントローラ障害時には正常コントローラに切り替え、処理を継続することにより信頼性を向上させる。また、ホストコントローラ側にバス切り替え用のミドルウェアを組み込むことなしにバス切り替えが可能である。このバス切り替え用ミドルは一般に高価であり、本発明は、システムとしての低コスト化にもつながる。

【図面の簡単な説明】

【図1】

本発明の1実施例を示す全体構成図である。

【図2】

キャッシュの内容の概念図である。

【図3】

コントローラ(又はホストI/F制御部3010)障害時の処理方式フロー図である。

【図 4】

コントローラ(又はホスト I/F 制御部 3010)障害時の処理方式フロー図である

【図 5】

コントローラ(又はホスト I/F 制御部 3010)復旧時の処理方式フロー図である

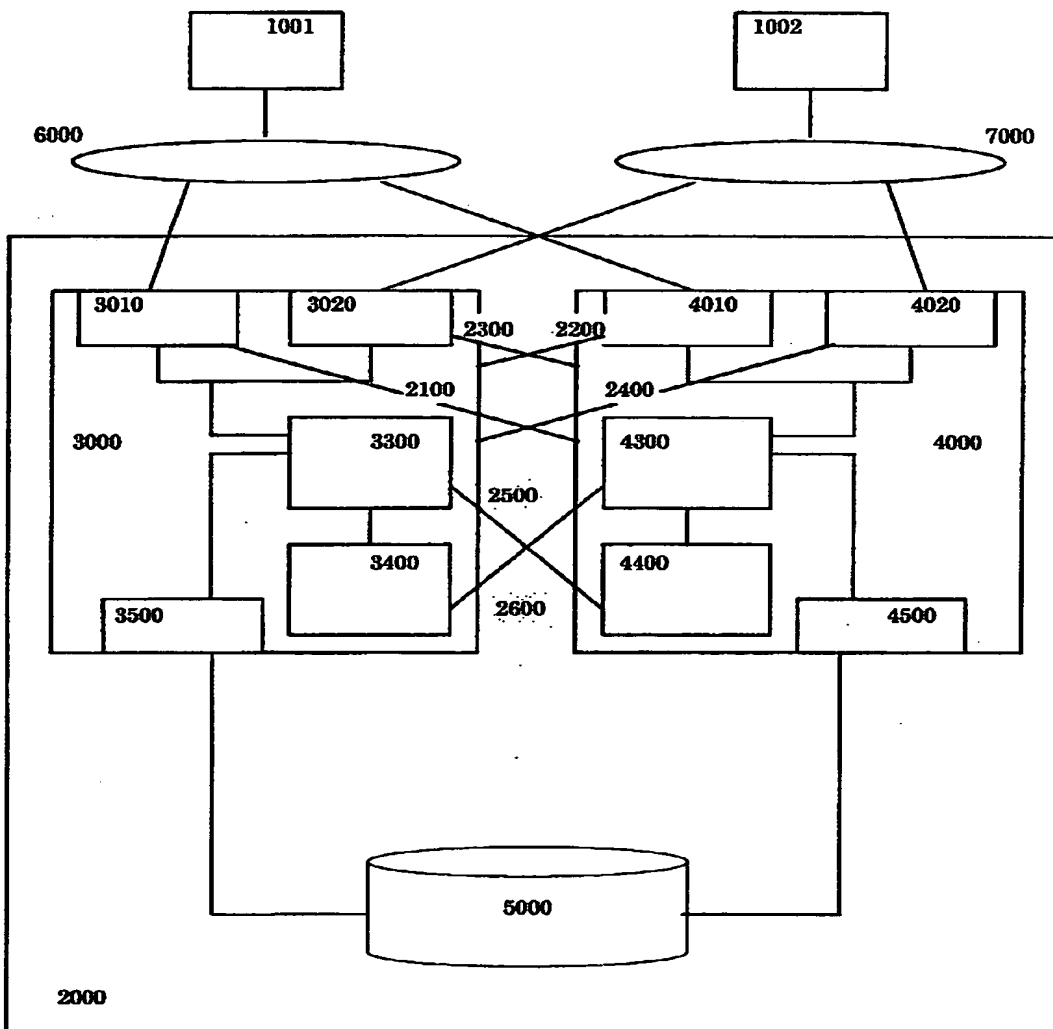
【符号の説明】

1001,1002…ホストコンピュータ、2000…ディスクアレイ装置、2100,2200,2300,2400…専用線、2500,2600…通信線、3000,4000…コントローラ、3010,3020,4010,4020…ホスト I/F 制御部、3300,4300…データ転送制御部、3400,4400…キャッシュ、3410,4410…データ部、3420,4420…コントローラ間通信エリア、3430,4430…I/F 情報エリア、3500,4500…DRV I/F 制御部、5000…ディスク装置群、6000,7000…Fibreチャネルループ。

【書類名】 図面

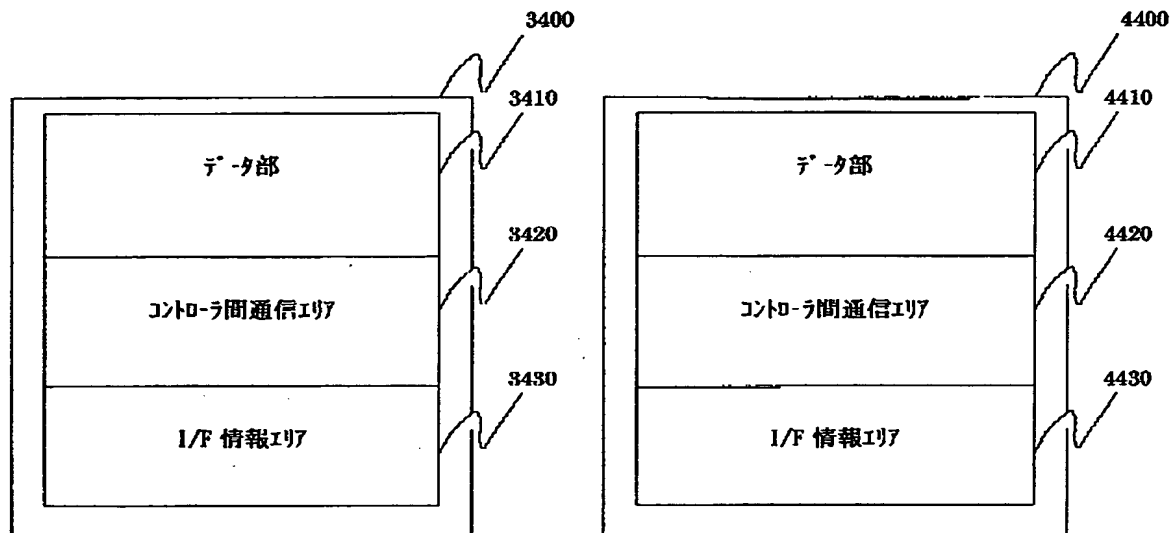
【図 1】

図 1



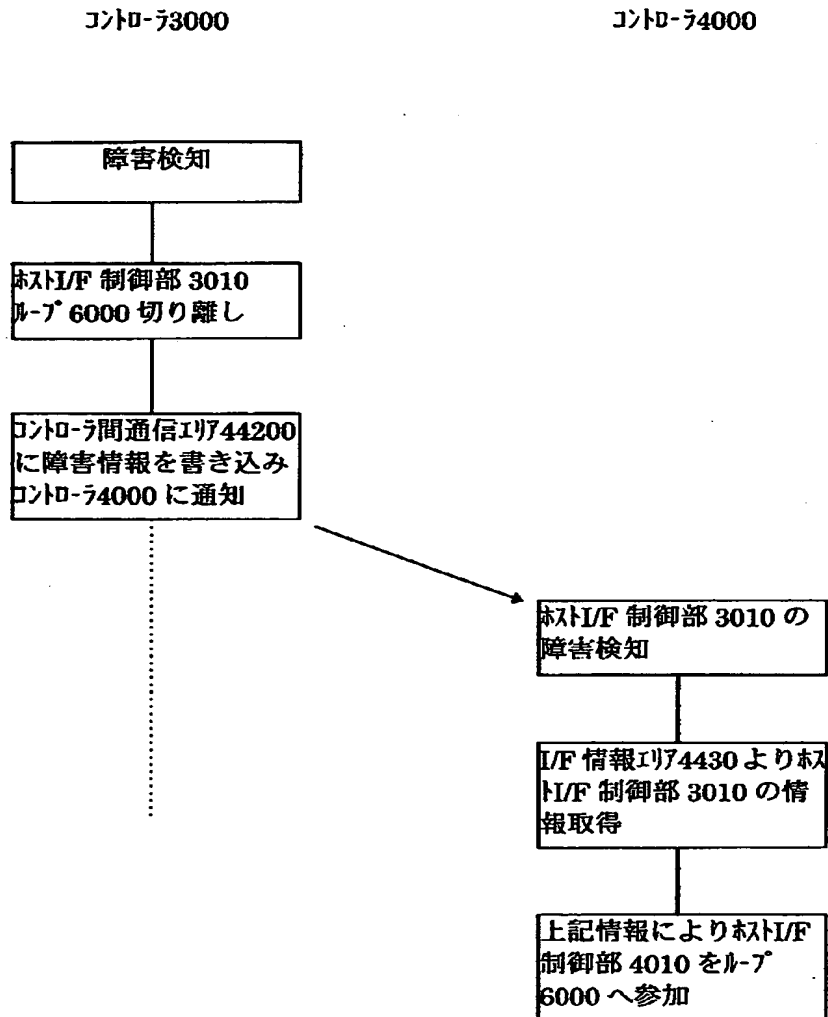
【図 2】

図 2



【図 3】

図 3

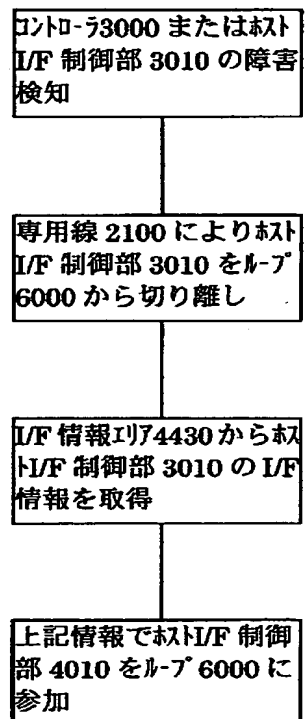


【図 4】

図 4

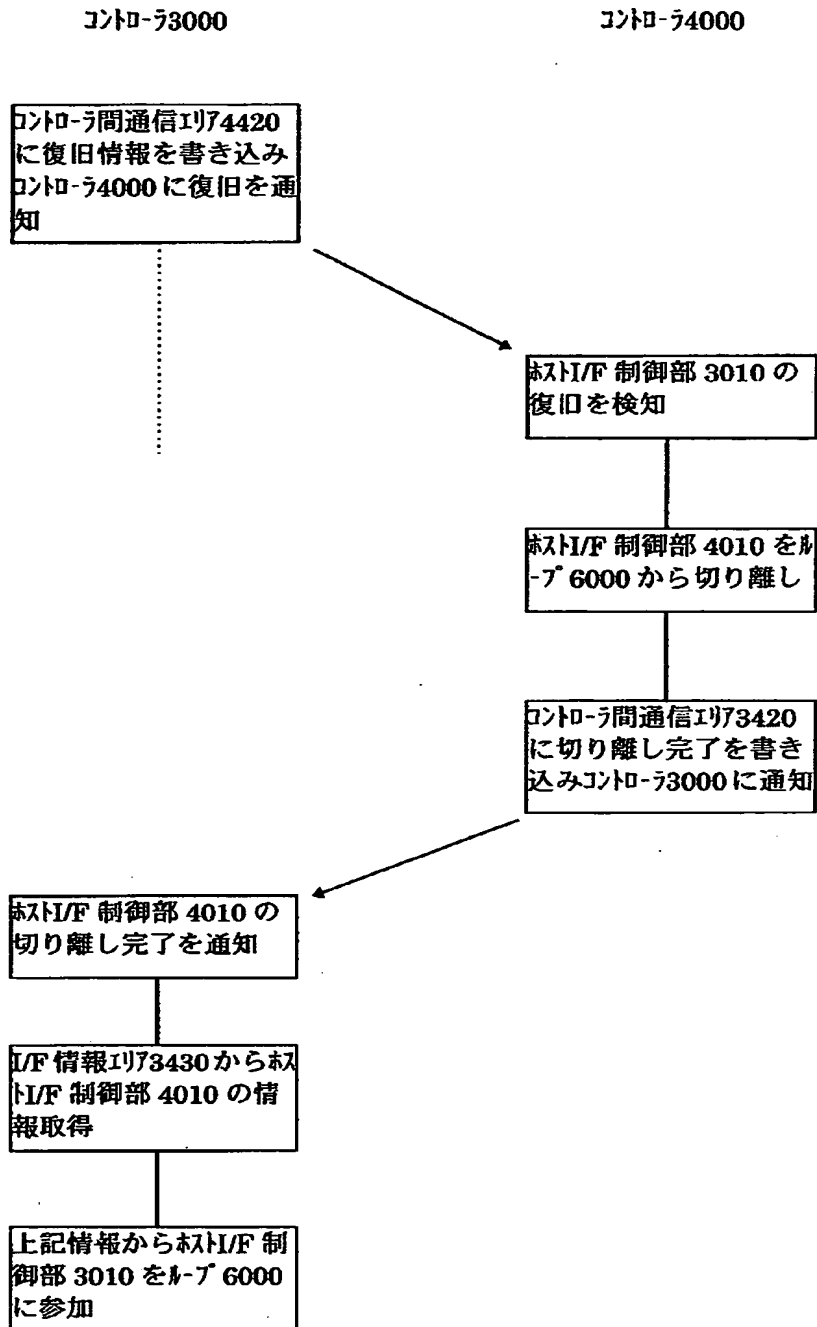
コントラ3000

コントラ4000



【図 5】

図 5





【書類名】 要約書

【要約】

【課題】

従来技術では、制御記憶装置側の機能により2重化したコントローラを現用系／待機系に分けて制御していた。そのため、正常時、構成要素の半分しか稼働できず性能を十分に引き出せずにいた。

【解決手段】

制御装置2000において、コントローラ3000,4000内の複数ポート3010,3020,4010,4020を各々、ポート毎に現用系、待機系を設定し、コントローラ1001或いはポート3010,3020の故障時には他のコントローラ1002内のポート4010,4020に交替する機能を持つことにより、正常時には、すべてのコントローラを稼働させることを可能とし高性能化を図る。

【選択図】 図1

出 願 人 履 歴 情 報

識別番号 [000005108]

1. 変更年月日	1990年 8月31日
[変更理由]	新規登録
住 所	東京都千代田区神田駿河台4丁目6番地
氏 名	株式会社日立製作所